

**KOREAN PATENT ABSTRACTS**

(11)Publication number: **2001-0092922 A**

(43)Date of publication of application: **27.10.2001**

(51)Int. Cl.

**G06F 17/30**

(21)Application number: **10-2000-0015595**

(22)Date of filing: **27.03.2000**

(71)Applicant: **LEE, HYUN KOO**

(72)Inventor: **LEE, HYUN KOO**

---

**(54) METHOD FOR SEARCHING SIMILAR WEB SITE**

(57) Abstract:

PURPOSE: A method for searching a similar web site is provided to enable a user to search a web site similar to the web site with the user was familiar by enabling a web server to register URLs of web sites and search words and compare a URL the user inputs with an arbitrary search word, thereby enabling the web server to estimate the degree of similarity.

CONSTITUTION: A user wants to search a web site similar to a specific web site. The user inputs a URL of the site. A web server stores search words including rankings and weight values and makes the first search word group of the search words. The user inputs an arbitrary search word to an input window. The search word the user inputs is compared with the first search word group. It is judged whether the search term is repeated. In case that there's the repeated search word, the search word in the first search word group is removed. The search word inputted from the user is inserted into a relevant ranking. The second search word group is formed. A candidate web site is selected. Weight values are given to search words in the second search word group.

(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(51) 。 Int. Cl. 7  
G06F 17/30

(11) 공개번호 특2001-0092922  
(43) 공개일자 2001년10월27일

(21) 출원번호 10-2000-0015595  
(22) 출원일자 2000년03월27일

(71) 출원인 이현구  
서울 강북구 번3동 226-1 한진아파트 102동 1901호  
(72) 발명자 이현구  
서울 강북구 번3동 226-1 한진아파트 102동 1901호  
(74) 대리인 박종현  
임영희

심사청구 : 있음

(54) 유사 웹사이트의 검색 방법

요약

본 발명은 유사 웹사이트를 검색하는 방법에 관한 것으로, 웹서버가 웹사이트들의 URL과 검색어들을 등록하고 사용자가 입력한 URL과 임의의 검색어를 비교하여 유사도를 평가함으로써 유사 웹사이트를 검색한다. 본 발명에 따른 방법은 단순히 검색어가 일치하는 빈도만을 측정하는 것이 아니라 일반적으로 인터넷 사용자들이 어느 웹사이트에 대해 기술하는 설명을 기초로 검색어들의 가중치를 설정하고, 이를 계속적으로 피드백하여 가중치를 갱신함으로써 실제 사용자들의 평가를 정확히 반영하여 유사 웹사이트 검색의 정확성을 높일 수 있는 장점이 있다.

대표도  
도 2

명세서

도면의 간단한 설명

도 1은 본 발명에 따른 등록된 웹사이트의 등록된 검색어와 가중치의 구조를 보인 도면,

도 2는 본 발명의 방법을 적용한 검색 웹사이트의 사용자 입력창의 구조를 보인 도면,

도 3a 및 도 3b는 본 발명의 방법에 따라 사용자가 원하는 유사 웹사이트를 검색하는 방법의 흐름을 보인 플로우차트이다.

## 발명의 상세한 설명

### 발명의 목적

#### 발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 유사 웹사이트를 검색하는 방법에 관한 것으로, 더욱 상세하게는 사용자가 특정 웹사이트의 URL(uniform resource locator)을 입력하고 원하는 임의의 검색어를 순위와 함께 입력함으로써, 검색어들의 출현 빈도와 가중치를 반영하여 검색 결과를 제시하는 유사 웹사이트의 검색 방법에 관한 것이다.

인터넷에는 전세계의 수많은 컴퓨터들이 연결되어 TCP/IP를 기반으로 통신이 이루어지며, 헤아릴 수 없을 정도로 많은 웹사이트들이 구축되어 다양한 콘텐츠를 제공하고 있다. 이들 웹사이트는 각자 나름대로 독특한 정보와 구성을 갖추고 있기 때문에, 이들 웹사이트들로부터 정치, 경제, 사회, 문화를 망라한 모든 분야에서 유용한 정보를 얻을 수도 있고 갖가지 콘텐츠를 이용할 수 있다.

따라서, 인터넷을 정보의 바다라고도 하며, 인터넷을 잘 이용하는 것이 개인의 발전이나 기업을 성패를 가름하는 것으로 인식되고 있는 실정에 있다. 더욱이 전자상거래를 통해 기존의 상거래 관행이 모두 혁신되고 인터넷을 기반으로 하는 금융 기법이 개발되어 온라인으로 모든 거래가 이루어지는 방향으로 발전하고 있다.

이와 같이 인터넷에 연결된 수많은 웹사이트들 중에서 사용자가 원하는 웹사이트를 찾아 접속하는 것이 쉽지 않다. 인터넷 사용자들을 위해 웹사이트를 찾는 것을 도와주는 웹사이트가 형성되어 서비스를 제공하고 있으며, 대표적으로 세계적으로는 www.yahoo.com, www.altavista.com, www.lycos.com 등이 유명하고, 국내에서는 www.simmani.com, www.naver.com, www.empas.com 등이 널리 이용되고 있다.

이들 검색 엔진(search engine)은 나름대로 독특한 방법으로 사용자가 입력하는 검색어를 기준으로 관련된 웹사이트를 찾아 목록을 제시한다. 예컨대, yahoo 검색엔진은 웹사이트를 직접 방문하여 전문가가 색인어를 결정하여 등록하고, 이 등록된 색인어와 사용자가 입력한 검색어를 비교하여 일치하는 웹사이트를 찾아주는 방법으로 실행된다.

또한, 최근의 발달된 서치엔진은 예컨대 "텔레비전을 살 수 있는 쇼핑물은 어디입니까?" 와 같은 자연어를 입력하도록 하여 자연어 검색 방식을 채용하는 것도 있다.

그러나, 이들 검색엔진의 방법들은 나름대로 장단점을 가지고 있어 경우에 따라서는 특정 검색엔진이 원하는 사이트를 바로 찾아주기도 하지만, 다수의 목록만을 제시할 뿐 이들 목록 자체가 너무 많은 웹사이트와 관련되어 있기 때문에, 도움이 되지 못하는 경우가 적지 않다. 또한 자신이 잘 알고 있는 웹사이트와 유사한 웹사이트를 찾고자 하는 경우에도 이들 종래의 서치엔진은 이러한 기능을 제공하지 않는다. 그러나 실제로 사용자들은 유명한 웹사이트를 잘 알고 있고, 그와 유사한 웹사이트를 비교하여 원하는 정보를 얻고자 하는 경우가 많다.

한편, 검색엔진의 성능은 사용자가 원하는 웹사이트 주소를 정확하게 제시하는 것으로 좌우되기 때문에, 그 검색 결과의 수가 너무 많은 경우에도 검색자는 그 검색 결과에 만족하지 못하게 되며, 특히 자신이 잘 알고 있는 웹사이트와 유사한 사이트를 찾는 경우에도 종래의 검색 엔진에 의한 일반적인 검색어 방식으로는 만족할만한 결과를 얻을 수 없었다.

#### 발명이 이루고자 하는 기술적 과제

따라서, 본 발명의 목적은 사용자 스스로 잘 알고 있는 웹사이트와 유사한 웹사이트를 효과적으로 검색해서 제시하는 인터넷을 통한 검색 방법을 제공하는 것이다.

## 발명의 구성 및 작용

이와 같은 본 발명의 목적은, 인터넷에 연결된 웹서버에 구축된 웹사이트에서 유사 웹사이트를 검색하는 방법에 있어서, 등록된 웹사이트의 URL 및 그 웹사이트를 대표하는 복수개의 검색어를 저장하는 단계와; 사용자가 홈페이지의 URL을 입력하는 단계와; 사용자가 임의의 검색어를 입력하는 단계와; 상기 사용자가 입력한 URL의 제 1 검색어 그룹에 대해 상기 사용자가 입력한 임의의 검색어를 추가하고, 검색어를 일정 개수로 제한한 제 2 검색어 그룹을 형성하는 단계와; 상기 제 2 검색어 그룹과 상기 웹서버에 등록된 웹사이트들의 등록된 검색어들을 비교하여, 일치하는 검색어의 개수가 많은 순서대로 일정 수의 검색 후보의 웹사이트를 선정하는 단계와; 상기 선정된 검색 후보의 웹사이트와의 등록된 검색어들과 상기 제 2 검색어 그룹을 비교하여, 사용자들의 이용도를 기준으로 유사도가 높은 순서대로 정렬한 상기 웹사이트들의 목록을 검색 결과로 제시하는 단계를 포함하는 방법이 제공된다.

본 발명의 방법은 인터넷에 연결된 서버에 구축된 검색 사이트로 구성된다. 즉, 사용자들은 인터넷을 통해 본 발명의 검색 사이트에 접속해서 원하는 검색 기능을 이용하게 된다. 따라서 본 발명의 검색 사이트는 독자적인 URL을 가지며, 사용자들은 이 URL을 이용해 인터넷을 통해 본 발명의 검색 사이트에 접속할 수 있다.

웹서버는 다수의 사용자를 수용할 수 있도록 고속의 전용선으로 인터넷에 연결되고 빠른 검색 결과를 제공할 수 있도록 컴퓨팅 파워가 충분한 서버를 사용하는 것이 바람직하다.

웹서버에는 웹사이트들이 등록되며, 이 등록된 웹사이트들이 검색 대상이 된다. 따라서, 웹서버는 가능한 한 인터넷상에 존재하는 모든 웹사이트 주소를 등록하여 저장하는 것이 바람직하다. 본 발명의 웹서버는 이들 등록된 웹사이트의 주소와 함께 해당 웹사이트에 대한 검색어를 함께 저장한다. 검색어는 사람, 특히 해당 분야의 전문가에 의해 검색어를 선정하여 등록할 수도 있다. 그러나 보다 바람직하게는 해당 웹사이트를 이용하는 사용자들의 사용 상태를 추적하여 검색어를 선정한다.

즉, 인터넷 사용자들이 스스로 선호하는 웹사이트 목록을 저장하는 온라인 북마크(bookmark)를 제공하는 웹사이트들이 다수 존재한다. 이러한 북마크를 온라인에 저장할 때, 사용자들은 북마크에 대해 간단한 설명을 기재하게 된다. 예컨대 한국방송공사의 "www.kbs.co.kr"에 대해서는 북마크하는 사람마다 다르기는 하겠으나 kbs, 방송, 언론, 한국, 기자, 뉴스, 드라마, 스포츠 등등의 설명을 붙인다. 따라서, 이렇게 실제 해당 웹사이트에 대해 사용자들이 사용하는 단어를 누적적으로 통계를 잡아서, 가장 많이 사용하는 단어의 순서대로 해당 웹사이트에 대한 검색어로 결정하여 등록한다. 이렇게 등록된 검색어는, 해당 웹사이트를 검색하고자 하는 검색자의 입력과도 대응할 가능성이 높기 때문에, 보다 정확한 검색이 이루어질 수 있는 것이다.

도 1은 이와 같이 웹사이트를 등록하면서 그 검색어를 등록한 예를 보인 것이다. 여기서 각 검색어에는 가중치가 설정된다. 이 가중치는 일반적으로 퍼센트의 개념으로 구성된다. 만약 "www.kbs.co.kr"이라는 웹사이트에 대해서 인터넷 사용자들의 북마크의 결과 이들의 검색어 순위와 사용빈도를 바탕으로 백분율을 구한 결과 "kbs"라는 단어가 35%로 1위, "방송"이라는 단어가 25%로 2위, "한국"이라는 단어가 10%로 3위.....였다면, 이러한 순위와 백분율에 따라 "www.kbs.co.kr"이라는 웹사이트는 도 1에 표시된 바와 같이 검색어(kbs, 방송, 한국, 언론, 뉴스.....)와 해당 검색어에 대한 가중치(3.5, 2.5, 1, 0.9, 0.7.....)가 등록된다. 도 1에서는 가중치의 합이 10으로 되는 경우를 설명하였으나, 합이 1로 되거나 합이 100으로 되는 방법으로 할 수도 있음은 물론이다.

마찬가지 방법으로, "www.kipo.go.kr"이라는 웹사이트에 대해 북마크 설명에 발견되는 빈도를 조사한 결과 "특허청"이라는 단어가 42%, "발명"이라는 단어가 23%, "특허"라는 단어가 11%, "실용신안"이라는 단어가 9%.....의 빈도로 나타났다면, 이 웹사이트는 도 1에 표시된 것과 같은 검색어와 그 가중치를 갖게 된다.

또한, 예컨대 "www.microsoft.com" 이라는 웹사이트에 대해서도 많은 사용자들의 북마크의 기술 결과 다양한 검색어가 등록될 수 있다. 여기서는 제 1 또는 제 2 검색어의 가중치가 위의 두 가지 예보다는 적은 것을 알 수 있다. 제 1 또는 제 2의 한두가지 검색어의 가중치가 높은 웹사이트는 그 단어에 의해서 그 웹사이트를 인식하는 사용자가 많다는 것으로, 이러한 유형의 웹사이트는 제 1 또는 제 2 검색어에 의해서 용이하게 그 웹사이트를 검색할 수 있다. 반대로, 검색어 사이에 가중치의 차이가 적은 웹사이트는 많은 사용자들이 여러 형태로 그 웹사이트를 인식하고 있는 것이어서, 검색어의 가중치가 더욱 중요한 의미를 갖게 된다.

이와 같이, 본 발명의 웹서버에는 웹사이트들의 URL과 그 URL의 검색어 및 그 검색어의 가중치가 동시에 저장된다. 하나의 URL에 검색어는 일반적으로 2이상의 복수개이며, 일반적으로 수십 개 이상의 다수의 검색어가 가중치와 함께 등록된다.

도 2에 도시된 바와 같이, 웹사이트들의 URL과 그 검색어 및 가중치는 본 발명의 웹서버에 데이터베이스(DB)의 형태로 저장된다. 또한, 웹서버에 구축된 본 발명의 방법에 따른 검색 웹사이트는 URL 입력창과 사용자가 임의로 검색어를 입력할 수 있는 검색어 입력창이 복수개 구비된다. 또한 각각의 임의의 검색어 입력창에는 순위를 지정할 수 있는 창이 별도로 제공되며, 사용자가 스스로 판단하여 검색어의 순위를 조절할 수 있도록 상향 버튼과 하향 버튼이 설치된다.

도 2에서 검색어 입력창은 5개의 검색어를 입력할 수 있도록 5개의 입력창이 도시되어 있고 순위는 초기 값으로 3으로 지정되어 있으나, 검색어 입력창은 필요에 따라 더 많이 또는 더 적게 제공될 수 있고 순위의 초기 값도 다르게 설정할 수 있다.

특정 웹사이트에 대해 그것과 유사한 웹사이트를 검색하고자 하는 사용자는 먼저 그 특정 웹사이트의 URL을 입력한다. 또한, 스스로의 판단으로 임의의 검색어를 1 또는 2 이상 복수개 입력하며, 각각의 검색어에 대해 원하는 순위를 지정한다.

이렇게 일정 수로 검색된 웹사이트에 대해 사용자가 입력한 특정 웹사이트와 가장 가까운 순서대로 후보 웹사이트의 순위를 재조정해야 한다.

이를 위해서 사용자에게 의해 입력된 URL의 등록된 검색어들(제 1 검색어 그룹)과 사용자가 입력한 임의의 검색어(제 1 검색어 내지 제 5 검색어)를 조합하여 새로운 검색어 그룹을 형성한다. 새로운 검색어 그룹의 형성은, 먼저 제 1 검색어 그룹의 검색어들이 가중치가 높은 순서대로 순위를 갖기 때문에, 사용자가 입력한 임의의 검색어와 그 순위를 존중하여 제 1 검색어 그룹의 순위로 임의의 검색어들을 삽입한다. 이렇게 삽입된 후에 그 순위에 있던 검색어들은 자연스럽게 다음 차순위로 자리 이동을 하며, 이 때 가중치는 삽입 또는 이동되는 검색어와는 무관하게 해당되는 순위에 고정된다.

이렇게 등록된 URL의 제 1 검색어 그룹과 사용자가 입력한 임의의 검색어들을 조합하여 형성된 새로운 검색어 그룹은 대체로 그 검색어의 수가 많기 때문에, 이를 일정한 수로 제한할 필요가 있다. 즉, 10 또는 15 순위 이하의 검색어들은 그 가중치가 매우 낮아서 실질적으로 검색의 결과에 대한 영향이 미미하기 때문이다.

따라서, 높은 순위부터 10, 15 또는 20개 정도로 제한하여 제 2 그룹의 검색어 그룹을 형성한다.

이렇게 형성된 제 2 검색어 그룹은 10개 내지 20개 정도의 검색어들을 포함하며, 이들 검색어를 기준으로 웹서버에 등록되어 있는 모든 웹사이트들의 등록된 검색어와 비교하여 일치하는 검색어가 많은 순서대로 검색 결과로 제시될 후보 웹사이트를 선정한다. 이 때 나열되는 후보 웹사이트의 수는 일정한 수, 예컨대 50개 내지 200개 정도로 제한할 수 있다.

선정된 후보 웹사이트들은 단순히 제 2 검색어 그룹과 일치하는 검색어가 많은 순서대로 결정된 것이기 때문에, 실제 사용자가 원하는 유사 웹사이트와 유사도가 반드시 일치하는 것은 아니다. 따라서, 검색자와 일반 사용자들의 이용 상태를 반영하여 유사도를 측정하여 보다 유사도가 높은 순서대로 후보 웹사이트들을 재배열하여야 한다.

이를 위해서 제 2 검색어 그룹과 후보 웹사이트들의 등록된 검색어를 비교하여 일치하는 검색어를 분리하고, 분리된 검색어들에 대해 제 2 검색어 그룹이 가지고 있는 가중치와 후보 웹사이트의 등록된 검색어의 가중치를 함께 곱한다. 이렇게 곱한 결과를 일치된 검색어마다 합해서 얻어진 합계가 해당 후보 웹사이트의 점수가 된다. 따라서, 후보 웹사이트들마다 이러한 방법으로 점수를 산출하여 점수가 높은 순서대로 정렬함으로써 검색 결과로 사용자에게 제시하게 된다.

이 때 제 2 검색어 그룹의 각각의 검색어에 대해 부여되는 가중치는 다음의 세가지 방법중 어느 하나로 결정한다.

먼저, 제 1 검색어 그룹은 사용자가 입력한 URL에 대해 웹서버가 가지고 있는 검색어로서 자체로서 가중치를 가지고 있다. 따라서, 사용자가 입력한 임의의 검색어들을 상술한 방법으로 제 1 그룹 검색어에 삽입하면 검색어들은 순위 변동을 하였으나, 각 순위에 대한 가중치들은 순위 변동 없이 그대로 유지되어 제 2 검색어 그룹의 각 순위에 대한 가중치로 한다.

둘째 방법으로, 웹서버는 등록된 모든 웹사이트들에 대해 검색어들과 그 검색어들의 가중치를 가지고 있다. 따라서, 이들 검색어들에 대해 각 순위별로 가중치를 산술 평균한 값을 구하여 이들 평균 값을 순위에 대한 일반적인 가중치로 하여 모든 제 2 검색어 그룹에 대해 적용하는 것이다.

첫째 방법은 해당 웹사이트에 대해 사용자들이 인식하는 가중치를 그대로 반영한 것이나, 그 웹사이트의 특수성에 따라 순위별 가중치의 편차가 심한 경우에는 그 가중치를 그대로 적용하는 것은 가중치의 왜곡을 가져올 수 있다. 둘째 방법은 일반적으로 검색어들의 검색어 순위에 대해 고정적인 가중치를 적용하는 것으로 첫째 방법과는 달리 해당 웹사이트의 특수성을 반영하지 못하는 단점이 있다.

셋째 방법은 첫째 방법으로 얻어지는 웹사이트 고유의 가중치와 둘째 방법으로 얻어지는 일반적으로 웹사이트들의 검색어들의 순위에 따른 가중치를 평균한 평균값을 다시 비례 평균하는 것이다. 이 때 비례 평균이라 함은 두 가중치를 더해서 2로 나눈 값으로 하거나 또는 어느 한쪽의 가중치를 좀더 존중하여 예컨대 고유 가중치에 2를 곱하여 평균 가중치와 더한 후 3으로 나누는 식으로 계산하는 것을 의미하는 것이다. 이 세 번째 방법은 해당 웹사이트의 특수성과 일반적인 검색어들의 순위를 모두 반영한 것으로 본 발명의 방법에 가장 바람직하다. 이 때 비례 평균을 위하여 어느 가중치를 존중하는 비례 계수를 적용하는가 하는 문제는 실제 검색 결과를 종합적으로 평가하여 실험적으로 확정할 수 있다.

이와 같이 세가지 방법 중에 어느 한 방법으로 제 2 검색어 그룹에 대해 가중치를 부여한 다음에, 상술한 후보 웹사이트들의 검색어들과 제 2 그룹 검색어들을 비교하여 동일한 검색어들을 추출한다. 추출된 검색어에 대해 제 2 검색어 그룹에서 부여한 가중치와 후보 웹사이트에서 부여된 가중치를 곱한 값을 구한다. 이렇게 동일한 검색어마다 가중치를 곱하고 검색어에 대해 모두 더한 값을 해당 웹사이트의 점수로 한다. 따라서, 점수가 높은 순서대로 후보 웹사이트들을 재배열함으로써 유사 웹사이트 검색의 결과로 제시한다.

본 발명의 방법을 도 3a와 도 3b를 참조하여 설명한다. 먼저 사용자는 원하는 특정 웹사이트와 유사한 웹사이트를 검색하기를 희망하고 있다. 따라서, 사용자는 그 URL을 입력한다(10).

본 발명에 웹서버는 가능한 모든 웹사이트들에 대해 온라인 북마크의 기술을 근거로 검색어들을 순위와 가중치를 포함하여 저장하고 있으므로, 그 등록되어 있는 검색어들을 제 1 검색어 그룹으로 한다(12).

사용자는 URL과 함께 스스로 중요하다고 생각되는 임의의 검색어를 입력창에 입력하며, 이 때 그 검색어의 순위도 함께 입력한다(14).

사용자가 입력한 검색어와 제 1 검색어 그룹을 비교하여 검색어가 중복되는 것이 있는지 판단한다(16). 중복되는 검색어가 없다면 제 1 그룹 검색어 그룹에 대해 임의의 검색어들은 사용자가 입력한 순위로 삽입되고, 그 삽입에 의해 해당 순위에 있던 검색어들은 자연히 차순위로 자리 이동을 한다(20).

만일 중복되는 검색어가 있다면, 동일한 검색어에 대해 이중으로 가중치를 부여하고 점수를 산출하는 것은 불합리하므로, 사용자가 입력한 검색어를 해당 순위로 삽입하고 제 1 검색어 그룹이 본래 가지고 있던 검색어는 삭제한다. 이렇게 함으로써 검색어의 중복을 회피할 수 있다(18, 20).

이런 방법으로 검색어 그룹을 형성되면 가중치가 낮아 검색 결과에 큰 영향을 미치지 않은 하위 순위 검색어들을 삭제한다. 따라서 가중치가 높은 순서에 따라 10 내지 20개 정도의 검색어로 제한한 제 2 검색어 그룹이 형성된다(22).

이 제 2 검색어 그룹을 기준으로 웹서버에 등록된 모든 웹사이트들의 검색어를 비교하여 일치하는 검색어의 수가 많은 순서대로 후보 웹사이트를 선정한다. 후보 웹사이트는 50 내지 200개, 바람직하게는 100개 정도로 제한한다(24).

제 2 검색어 그룹은 검색어의 그 순위만 가지고 있고 가중치는 없으나 상술한 세 가지 방법 중에서 어느 한 방법으로 제 2 검색어들에 대해 가중치를 부여한다(26).

가중치가 부여되면 제 2 검색어 그룹과 후보 웹사이트들의 검색어를 비교하여 동일한 검색어들을 찾아낸다(28).

동일한 검색어에 대해 제 2 검색어 그룹이 가지고 있는 가중치와 후보 웹사이트가 가지고 있는 가중치를 곱한 값을 구한다. 얻어진 검색어에 대해 이와 같이 양쪽의 가중치를 곱한 값을 모두 더한 값을 그 웹사이트의 점수로 한다(30).

따라서, 후보 웹사이트들을 점수가 높은 순서대로 재배열함으로써 본 발명의 방법에 따른 유사 웹사이트의 결과로서 제시된다.

본 발명의 검색 방법을 예를 들어 설명하기로 한다.

먼저, 사용자가 " www.microsoft.com" 과 유사한 웹사이트를 검색하기를 원하는 경우에, 사용자는 본 발명의 웹사이트에 접속해서 URL 입력창에 " www.microsoft.com" 을 입력하고, 사용자 스스로 임의의 검색어로서 리눅스(순위 1), 공룡(순위 4), OS(순위 7), 대통령(순위 9)을 입력하고 검색을 실행하였다고 가정하자. 따라서 웹서버에 등록된 제 1 검색어 그룹과 여기에 사용자가 입력한 임의의 검색어를 추가한 새로운 검색어 그룹은 다음과 같다.

제 1 검색어 그룹: ① 윈도우즈(2.2), ② 빌게이츠(2.1), ③ OS(1.9), ④ 소프트웨어(1.8), ⑤ 소송(1.1), ⑥ 독점((0.9), ⑦ 리눅스(0.8), ⑧ 벤처(0.7), ⑨ 아이콘(0.6), ⑩ 미국(0.5), ⑪ 생각속도(0.4), ⑫ 부자(0.3), ⑬ 공룡(0.2), ⑭ 꿈(0.1).....

새로운 검색어 그룹: ① 리눅스(2.2), ② 윈도우즈(2.1), ③ 빌게이츠(1.9), ④ 공룡(1.8), ⑤ 소프트웨어(1.1), ⑥ 소송((0.9), ⑦ OS(0.8), ⑧ 독점(0.7), ⑨ 대통령(0.6), ⑩ 벤처(0.5), ⑪ 아이콘(0.4), ⑫ 미국(0.3), ⑬ 생각속도(0.2), ⑭ 부자(0.1) ⑮ 꿈(0.09).....

여기서 보는 바와 같이 사용자가 입력한 임의의 검색어는 입력한 순위로 삽입되고 그 순위에 있던 검색어는 차순위로 이동한다. 사용자가 입력한 검색어가 제 1 검색어 그룹에 존재하는 경우(리눅스, 공룡, OS)에는 그들 단어가 해당 순위로 이동하고, 존재하지 않는 경우(대통령)에는 입력한 순위(9)로 그대로 삽입된다. 이와 같이 검색어들이 자리 이동 또는 삽입되는 경우에도 제 1 검색어 그룹에 등록되어 있는 가중치는 전혀 이동하지 아니하고, 그 순위로 고정되어 있다 (첫번째 가중치 설정 방법을 적용).

형성된 새로운 검색어 그룹은 그 수가 많으므로, 이를 일정 수로 제한한다. 예컨대 검색어를 10개로 제한하였다면, 제 2 그룹 검색어 그룹은 다음과 같다.

제 2 검색어 그룹: ① 리눅스, ② 윈도우, ③ 빌게이츠, ④ 공룡, ⑤ 소프트웨어, ⑥ 소송, ⑦ OS, ⑧ 독점, ⑨ 대통령, ⑩ 벤처

10개의 검색어로 제한된 제 2 검색어 그룹으로 이들 검색어와 일치하는 검색어를 많이 가지고 있는 후보 웹사이트를 검색하여, 일치하는 검색어의 수가 많은 순서대로 후보 웹사이트를 일정 수, 예컨대 100개를 선정한다.

후보 웹사이트 100개가 다음과 같이 검색되었다고 가정하자.

(1) www.xxx.com: ① 리눅스(2.4), ② 전자화폐(1.9), ③ OS(1.1), ④ 스크린(1), ⑤ 다운로드(0.9), ⑥ 독점(0.8), ⑦ 쇼핑물(0.7), ⑧ 도스(0.6), ⑨ 쇼핑카드(0.5), ⑩ 신용카드(0.4), ⑪ 소프트웨어(0.3), ⑫ 벤처(0.2), ⑬ 윈도우(0.1), ⑭ 꽃(0.09), ⑮ 생일(0.08)....

(2) www.yyy.com: ① 생각속도(3.1), ② 그림(2.1), ③ 윈도우(1.9), ④ 택배(1.1), ⑤ 리눅스(0.9), ⑥ 빌게이츠(0.8), ⑦ 엽서(0.7), ⑧ 소송(0.5), ⑨ 경매(0.4), ⑩ 역경매(0.3), ⑪ 고문서(0.2), ⑫ 경품(0.1), ⑬ 신문(0.1), ⑭ 잡지(0.1).....

(3) www.zzz.com: ① 리눅스(4.2), ② 빌게이츠(2.1), ③ 윈도우(1.9), ④ 페인트샵(0.9), ⑤ 포토샵(0.8), ⑥ 다운로드(0.7), ⑦ 신용카드(0.7), ⑧ 업그레이드(0.6), ⑨ 웨어웨어(0.5), ⑩ 프리웨어(0.5), ⑪ 업로드(0.4), ⑫ 게임(0.4), ⑬ 스타크래프트(0.3), ⑭ 삼국지(0.3).....

(4) www.....

(100) www.....

여기서, 제 2 검색어 그룹에 대해 가중치를 부여한다. 그 가중치는 상술한 첫 번째 "www.microsoft.com"의 순위별 고유 가중치를 그대로 적용한다(첫번째 해당 웹사이트의 순위별 고유 가중치를 그대로 적용).

이 때 유사도를 평가하는 방법으로 첫 번째 www.xxx.com의 검색어들[① 리눅스(2.4), ② 전자화폐(1.9), ③ OS(1.1), ④ 스크린(1), ⑤ 다운로드(0.9), ⑥ 독점(0.8), ⑦ 쇼핑물(0.7), ⑧ 도스(0.6), ⑨ 쇼핑카드(0.5), ⑩ 신용카드(0.4), ⑪ 소프트웨어(0.3), ⑫ 벤처(0.2), ⑬ 윈도우(0.1), ⑭ 꽃(0.09), ⑮ 생일(0.08)....]과 상기 제 2 검색어 그룹[① 리눅스(2.2), ② 윈도우(2.1), ③ 빌게이츠(1.9), ④ 공룡(1.8), ⑤ 소프트웨어(1.1), ⑥ 소송(0.9), ⑦ OS(0.8), ⑧ 독점(0.7), ⑨ 대통령(0.6), ⑩ 벤처(0.5)]을 비교하여 공통된 검색어를 찾는다. 여기서는 리눅스, OS, 독점, 소프트웨어, 벤처, 윈도우의 6개 검색어가 일치하였다. 이와 같이 공통된 검색어를 분리한 후에 이들 검색어에 대해 제 2 검색어 그룹에서 부여된 가중치와 www.xxx.com의 등록된 가중치를 곱한 후에, 검색어 마다 그 곱한 값을 합하여 합계를 구한 것을 www.xxx.com의 점수로 한다. 따라서, www.xxx.com의 점수의 계산은 다음과 같다.

(1) www.xxx.com: 리눅스( $2.4 \times 2.2$ ) + OS( $1.1 \times 0.8$ ) + 독점( $0.8 \times 0.7$ ) + 소프트웨어( $0.3 \times 1.1$ ) + 벤처( $0.2 \times 0.5$ ) + 윈도우( $0.1 \times 1.1$ ) = 7.26

또한 두 번째 www.yyy.com은 제 2 검색어 그룹과 윈도우, 리눅스, 빌게이츠 및 소송의 4개 검색어가 일치하였다. 따라서, 이 웹사이트의 점수는 다음과 같다.

(2) www.yyy.com: 윈도우( $1.9 \times 2.2$ ) + 리눅스( $0.9 \times 2.2$ ) + 빌게이츠( $0.8 \times 1.9$ ) + 소송( $0.5 \times 0.9$ ) = 8.13

www.zzz.com은 제 2 검색어 그룹과 리눅스, 빌게이츠 및 윈도우의 3개 검색어가 일치하였다. 따라서, 이 웹사이트의 점수는 다음과 같다.



(3) www.zzz.com: 리눅스( $4.2 \times 2.4$ ) + 빌게이츠( $2.1 \times 1.9$ ) + 윈도우( $1.9 \times 0.1$ ) = 14.26

이와 같이 후보 웹사이트들에 대해 같은 방법으로 일치된 검색어를 분리하여 당해 웹사이트의 가중치와 제 2 검색어 그룹에서 부여된 가중치를 곱한 후에 이를 합제한 점수를 구한 결과, 최고 점수로부터 순위가 14.26, 8.13, 7.26.....이 었다면, 이 검색의 결과 목록은 다음과 같이 표시된다.

(1) www.zzz.com

(2) www.yyy.com

(3) www.xxx.com

.....

이 목록이 검색의 결과로서 검색을 실행한 사용자에게 제시된다.

앞에서 첫 번째 웹사이트의 순위별 고유 가중치를 제 2 그룹 검색어에 그대로 적용하는 방법으로 유사 웹사이트를 검색 하였으나, 만일 상술한 두 번째 방법을 선택한다면, 그 가중치는 어떠한 검색에서도 동일하게 검색어들의 순위별 고정 가중치, 예컨대 제 1 검색어는 3.2, 제 2 검색어는 2.8, 제 3 검색어는 1.9, 제 4 검색어는 1.1.....와 같은 가중치를 갖 는다. 이러한 가중치가 부여된 검색어를 사용하여 후보 웹사이트들의 점수를 산출한다면 후보 웹사이트들의 검색 결과 는 첫 번째 가중치를 사용한 결과와 달라질 수 있다.

또한 세 번째 가중치 부여 방법을 제 2 그룹 검색어에 대해 적용하되 단순히 양쪽 가중치를 더하여 2로 나눈 값을 부여 한다면, 제 1 검색어에는  $(3.2+2.2)/2=2.7$ , 제 2 검색어에는  $(2.8+2.1)/2=2.45$ , 제 3 검색어에는  $(1.9+1.9)/2=1.9$ , 제 4 검색어에는  $(1.1+1.8)/2=1.45$ ....와 같은 가중치가 부여된다. 이렇게 부여된 가중치를 사용하여 후보 웹사 이트의 점수를 계산하면 또 다른 결과를 출력하게 된다.

본 발명의 방법에서, 웹사이트의 등록은 온라인 북마크를 이용하는 이용자들의 이용상태를 추적하여 데이터베이스를 추 가 및 갱신하는 방법을 사용하며, 필요한 경우에 본 발명의 웹서버 운영자에 의해 전문가 집단을 이용해 등록을 할 수 있다.

또한, 등록된 웹사이트의 가중치를 정하는 방법으로 온라인 북마크의 사용자 설명을 기초로 이를 누적하여 백분율을 구 하였으나, 온라인 북마크 이외에 본 발명에 따른 검색 방법을 이용하여 검색을 실행하는 사용자들이 URL을 입력하고 그것에 대해 추가로 입력하는 임의의 검색어를 입력으로 하여 가중치에 반영하는 것이 바람직하다. 이렇게 계속해서 등 록된 검색어들의 가중치를 사용자들의 사용 상황에 비추어 갱신함으로써 더욱 정확한 검색 결과를 얻을 수 있는 것이다.

#### 발명의 효과

본 발명에 따른 유사 웹사이트 검색방법은 사용자가 원하는 웹사이트와 가장 유사한 웹사이트를 검색하여 그 목록을 제 시한다. 본 발명은 웹을 이용하는 사용자들의 검색어에 대한 평가를 반영한 가중치 및 검색자 자신이 입력한 검색어를 조합하여 적용함으로써 사용자가 원하는 유사 웹사이트를 효과적으로 검색하여 결과를 출력할 수 있다. 또한, 계속해서 등록된 웹사이트의 검색어 가중치를 웹 사용자들의 사용 형태를 반영하여 갱신함으로써 본 발명의 방법에 따른 검색 결 과의 질이 향상될 수 있는 것이다.

이상에서 바람직한 실시예에 의거하여 본 발명을 상세히 설명하였으나, 본 발명의 사상과 첨부된 특허청구범위 내에서 이 분야의 전문가에게 자명한 치환, 변경 또는 변형은 본 발명의 범위에 속하는 것으로 이해되어야 한다.

.(57) 청구의 범위

청구항 1.

인터넷에 연결된 웹서버에 구축된 웹사이트에서 유사 웹사이트를 검색하는 방법에 있어서,

등록된 웹사이트의 URL 및 그 웹사이트를 대표하는 복수개의 검색어를 저장하는 단계와;

사용자가 홈페이지의 URL을 입력하는 단계와;

사용자가 임의의 검색어를 입력하는 단계와;

상기 사용자가 입력한 URL의 제 1 검색어 그룹에 대해 상기 사용자가 입력한 임의의 검색어를 추가하고, 검색어를 일정 개수로 제한한 제 2 검색어 그룹을 형성하는 단계와;

상기 제 2 검색어 그룹과 상기 웹서버에 등록된 웹사이트들의 등록된 검색어들을 비교하여, 일치하는 검색어의 개수가 많은 순서대로 일정 수의 검색 후보의 웹사이트를 선정하는 단계와;

상기 선정된 검색 후보의 웹사이트와의 등록된 검색어들과 상기 제 2 검색어 그룹을 비교하여, 사용자들의 이용도를 기준으로 유사도가 높은 순서대로 정렬한 상기 웹사이트들의 목록을 검색 결과로 제시하는 단계를 포함하는 방법.

청구항 2.

제 1 항에 있어서, 상기 등록된 웹사이트의 검색어마다 가중치를 부여하고, 그 검색어의 선정과 가중치는 그 웹사이트를 이용하는 사용자들의 이용도를 평가하여 결정하는 것을 특징으로 하는 방법.

청구항 3.

제 2 항에 있어서, 상기 사용자들의 이용도는, 상기 등록된 웹사이트에 대한 사용자들의 온라인 북마크의 설명을 기준으로 결정되는 것을 특징으로 하는 방법.

청구항 4.

제 2 항에 있어서, 상기 사용자는 임의의 검색어를 원하는 순위를 부여하여 입력하고, 임의의 검색어들은 부여된 순위로 제 1 검색어 그룹에 추가되고, 상기 추가에 의해 제 1 검색어 그룹의 검색어들은 차순위로 순위 변동하며, 추가된 임의의 검색어가 제 1 검색어 그룹에 이미 존재하는 경우에는 동일 검색어의 중복을 피하도록 상기 제 1 검색어 그룹에 이미 존재하던 검색어가 삭제되는 것을 특징으로 하는 방법.

청구항 5.

제 4 항에 있어서, 상기 제 2 검색어 그룹의 각 검색어들에 대한 가중치는, 상기 등록된 검색어들의 일반적인 각 순위별 평균 가중치를 부여하는 것을 특징으로 하는 방법.

청구항 6.

제 4 항에 있어서, 상기 제 2 검색어 그룹의 각 검색어들에 대한 가중치는, 제 1 검색어 그룹에 부여된 순위별 고유의 가중치를 그대로 유지하는 것을 특징으로 하는 방법.

청구항 7.

- 제 4 항에 있어서, 상기 제 2 검색어 그룹의 각 검색어들에 대한 가중치는, 상기 등록된 검색어들의 일반적인 각 순위별 평균 가중치와 제 1 검색어 그룹에 부여된 순위별 고유의 가중치를 비례 평균한 값을 가중치로 부여하는 것을 특징으로 하는 방법.

청구항 8.

제 5 항 내지 제 7 항중 어느 한 항에 있어서, 상기 사용자들의 이용도를 기준으로 하는 유사도의 평가는, 상기 제 2 검색어 그룹과 상기 후보 웹사이트들의 등록된 검색어를 비교하여 동일한 검색어를 추출하고, 추출된 검색어마다 해당 검색어에 대해 상기 제 2 검색어 그룹에서 부여된 가중치와 상기 후보 웹사이트의 등록된 검색어에 부여된 가중치를 곱한 값을 구하고, 추출된 검색어마다 상기 곱한 값을 얻어서, 이들 곱한 값을 모두 합한 합계를 해당 후보 웹사이트의 점수로 하고, 그 점수가 높은 순서대로 후보 웹사이트를 배열하여 목록을 작성하는 것을 특징으로 하는 방법.

청구항 9.

제 2 항에 있어서, 상기 등록된 웹사이트에 대한 등록된 검색어와 가중치에 대해 상기 검색을 실행하는 사용자들이 입력한 임의의 검색어와 가중치를 피드백하여 상기 등록된 검색어와 가중치를 변경하는 것을 특징으로 하는 방법.

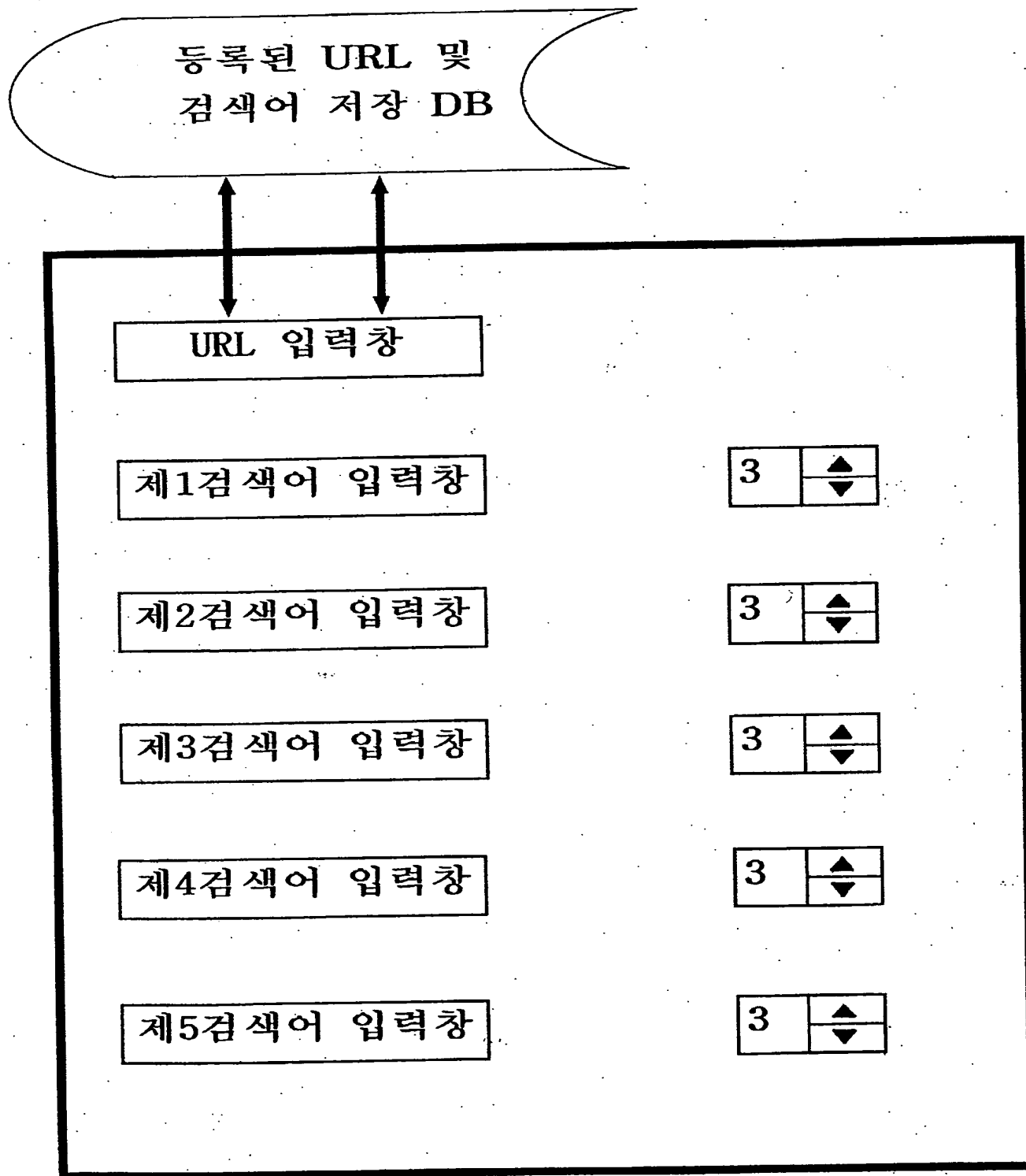
도면

도면 1

www.kbs.co.kr	kbs	3.5	방송	2.5	한국	1	인론	0.9	뉴스	0.7	드라마	0.6	스포츠	0.5
	음악회	0.5	연예인	0.4	기자	0.4	mbc	0.3	전파	0.2	NHK	0.1	.....	...
www.kipo.go.kr	특허청	4.2	발명	2.3	특허	1.1	실용신안	0.9	상표	0.4	의장	0.2	서비스표	0.1
	심판	0.1	심사	0.1	소송	0.1	심사관	0.1	변리사	0.1	대전	0.1	.....	...
www.microsoft.com	윈도즈	2.2	빌게이츠	2.1	OS	1.9	소프트	1.8	소송	1.1	독점	0.9	리눅스	0.8
	벤처	0.7	아이콘	0.6	미국	0.5	생각속도	0.4	부자	0.3	공룡	0.2	꿈	0.1
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.

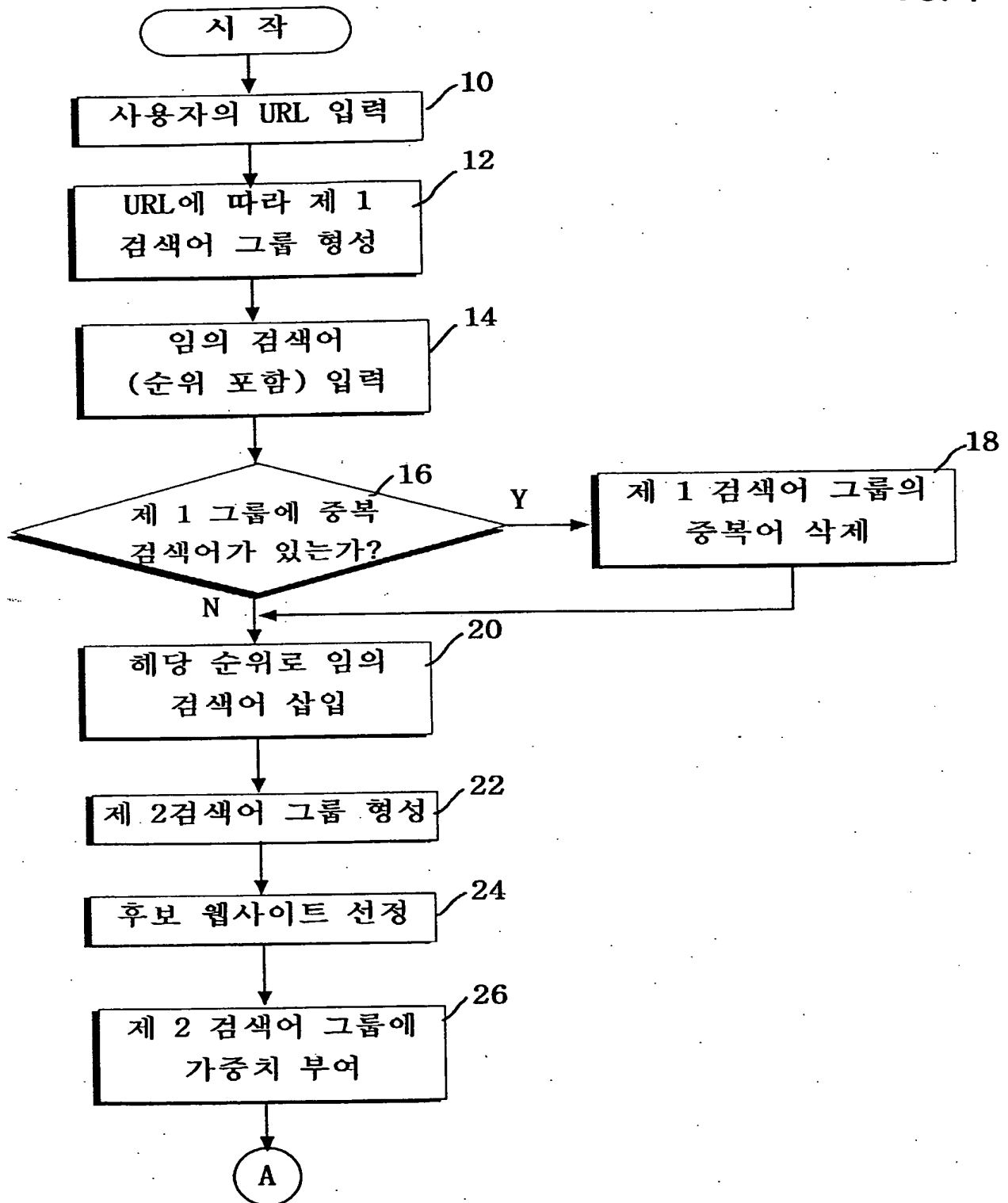
도면 2

BEST AVAILABLE COPY



도면 3a

BEST AVAILABLE COPY



도면 3b

BEST AVAILABLE COPY

